

**자료 입력을 위한 편집**

# 자료 입력을 위한 클리닝 과정

## 분석 자료 클리닝 과정



# 무응답의 유형

- 무응답 요인
  - 부재
  - 거부
  - 표본 부적격
- 무응답이 조사에 미치는 영향
  - 분포
  - 추정 및 추론 결과 왜곡
  - 가중치에 영향을 제공
  - 분산의 증가
- 무응답의 유형
  - 단위무응답
  - 항목무응답

# 무응답 문항 검토 및 수정

- 1) 응답해야 할 모든 문항의 응답 여부를 확인
  - 단위무응답(unit nonresponse) :
    - (1) 표본 비접촉/부재 : 표본 적격 미지
    - (2) 부적격 표본 : 표집틀의 포함 오류
    - (3) 조사 참여 거부 : 교체 가능
  - 항목무응답(item nonresponse)

## 2) 무응답 문항 발견 시

- 기본적으로 재조사가 원칙
- 무응답 처리 방안

### (1) 그대로 두는 경우

- 자료 입력 시 9, 99, 999 등 입력 자리 숫자의 최대값으로 부여
- 무응답 코딩은 모수 추정 시 제외

### (2) 대체(imputation)

- 평균대체
- 회귀대체
- Deck : Cold deck, Hot deck
- 다중대체
- 주의 : 대체 자료는 대체여부를 하나의 변수로 처리해 분산 추정 시 제외

# 개방형 코딩

1) 코딩(coding)이란? 응답자가 직접 기록한 응답에 숫자나 부호를 부여하는 과정

1) 예 : 남자는 1, 여자는 0

2) 색깔 : 노랑 1, 빨강 2, 청색 3, 하늘색 4, 검정 5 등

2) 선택형에서의 코딩

– 선택형으로 부여된 숫자 이후의 번호를 순서대로 부여

3) 개방형에서의 코딩

– 설문지를 보면서 입력 번호를 부여

– 여러 명이 코딩작업을 할 때 통일된 번호로 코딩하기 위한 관리가 요구됨

# 지분형 문항의 검토 및 수정

예)

문22 현재 재택전자(인터넷, PC통신) 민원처리제가 시행되고 있다는 사실을 알고 계십니까?

① 예                      ② 아니오 → 문 23으로 가시오

문22-1. 그러면 재택전자 민원처리제에 대해 어느 정도 만족하셨습니다까?

전혀 불만족	대체로 불만족	보통	대체로 만족	매우 만족
-----	-----	-----	-----	-----
①	②	③	④	⑤

잘 모름 / 무응답  
☐  
⑥

- 1) 지분을 위한 문항의 응답과 지분형 문항의 대상자가 일치하는지 검토
  - 지분을 위한 문항의 응답 기준으로 처리
  - 참고 : 빈도분석으로 검토하는 경우
    - (1) 지분 기준 문항 & 지분형 무응답/결측 확인
    - (2) (2) 지분형 응답 여부 & 지분 기준 문항 확인
- 2) 무응답과 결측(missing value)
  - 무응답 : 응답을 해야 하는데 응답하지 않은 것으로 입력 자료 부여(9, 99 등)
  - 결측 : 응답 대상이 아니므로 응답하지 않은 것으로 입력 자료 없음(결측 처리)

# 신뢰성 및 일치성 문항 검토 및 수정

## 1) 신뢰성 문항 검토

- 응답자가 문항에 거짓없이 응답(측정오류 확인)하는지 확인하기 위해 작성된 문항으로 설문지내 서로 유사한 문항을 작성해 확인하는 문항
- 유사 문항에 대한 응답이 다른 경우, 응답자가 설문 문항에 전반에 대해 거짓 혹은 편향되게 응답한 것으로 판단해 분석자료에서 제외하는 것이 바람직

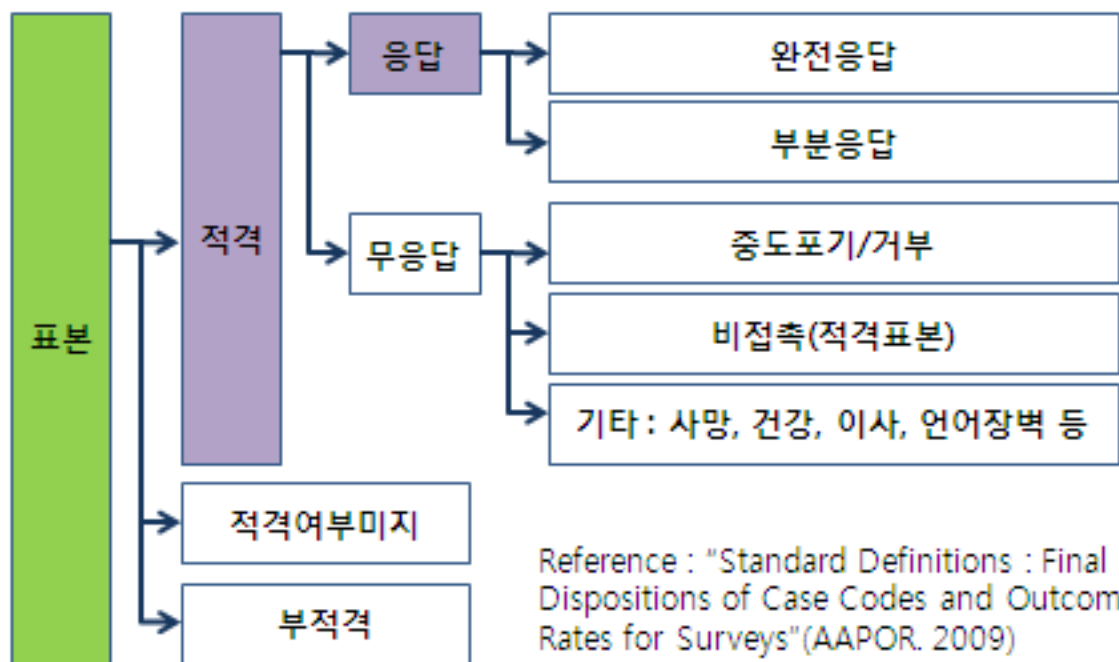
## 2) 일치성 문항 검토 및 수정

- 응답자가 수치 자료로 응답하는 경우, 수치 자료의 일치성을 검토해 수정하는 과정
- 수정 기준 : 응답자의 응답자료를 기준으로 판단하는 것이 바람직
  - (1) 응답 자료를 기준으로 합을 수정하는 방법
  - (2) 합을 기준으로 응답 자료를 수정하는 방법



# 분석 설문지 분류/유효 분석 표본의 결정

## 표본의 응답 결과 유형



Reference : "Standard Definitions : Final Dispositions of Case Codes and Outcome Rates for Surveys" (AAPOR, 2009)  
- 김서영 외(2010) 참조 인용

# 1) 분석 제외 자료의 분류 기준

- [참고] **응답에 대한 표준 정의** (AAPOR, 2009) : 전화, 면접, 우편조사 적용

- ① 반드시 응답해야 할 문항 중 50% 미만(중도포기), 50~80% 미만(부분응답), 80% 이상 (완전응답)으로 간주
- ② 전체 문항 중 50% 미만(중도포기), 50~80% 미만(부분응답), 80% 이상 (완전응답)으로 간주
- ③ 핵심 질문 중 응답한 항목이 전체 핵심 질문의 50% 미만(중도포기), 50~99% 이하 (부분응답), 100% (완전응답)으로 간주
- ④ 위의 세 가지를 조합하여 정의 가능 : 핵심질문은 완전응답, 그 외 질문은 80% 이상 응답만을 완전응답으로 간주하는 등

## 2) 응답 제외 설문지 기준

- (1) 일반적으로 응답 문항 전체 중 90% 미만인 설문지
- (2) 분석을 위한 기본 문항은 모두 응답하고 나머지 문항 중 80% 미만인 설문지
- (3) 신뢰성 문항에서 신뢰성에 의심이 가는 설문지

## 3) 응답 제외 설문지의 처리

- 분석 제외 자료는 단위 무응답 자료로 간주해 무응답 조정 가중치 부여
- 보고서 작성 시 분석 제외 자료 현황을 제시
  - (1) 전체 수집 자료 크기
  - (2) 분석 제외 자료 크기 : 제외 사유 표시
  - (3) 유효 분석 자료 크기

# 자료 입력

## 1) 문항별로 변수명을 부여

- 변수명은 분석하는 통계 패키지의 변수 이름 부여 규칙에 따라 부여
- 데이터셋에 변수를 입력할 때 하나의 답은 하나의 변수로 입력하는 것이 원칙
  - 하나의 변수에 답에 2개 존재할 수 없음
- 예 : [문항 27] 다음에서 두 가지를 선택하십시오.
  - 첫 번째 답 : q27\_1
  - 두 번째 답 : q27\_2

## 2) 입력 방법

- (1) 간접 입력 방법 : 엑셀, 메모장, 한글 등
- (2) 직접 입력 방법 : 분석 프로그램의 자료 입력 창에서 입력

# 빈도분석을 이용한 클리닝

## 1) 자료 클리닝에서의 빈도분석

- 목적 :

- (1) 응답 및 입력 자료의 오류 수정을 위한 기초 분석
- (2) 수치 자료의 일치성 검토 및 확인을 위한 기초 분석
- (3) 연속형 자료의 범주화를 위한 기초 분석

## 2) 응답 및 입력 오류 자료의 검토 및 수정

- 모든 문항에 대한 빈도분석을 수행
- 문항별로 응답 범주를 벗어난 이상값 검토 및 수정

## 3) 일치성 검토 및 수정

- 수치 자료에 대한 일치성을 확인하기 위해 개별 자료의 합과 응답된 합과 일치 여부를 확인
- 수정 과정은 입력 전 일치성 검토 및 수정 과정과 동일

## 4) 연속형 수치 자료에 대한 범주화 변환

### – 범주화 변환 방법

- (1) 연속형 수치 자료의 범주화 작업은 통계적 이론을 근거로 범주화 시도
- (2) 사회적 관행에 따른 범주화

### – 범주화 변환 시 주의 사항

- (1) 하나의 값이 두 개 범주에 속하지 않도록 주의 :
  - 미만, 이하/이상 사용 주의
- (2) 범주화한 변수명을 원자료의 변수와 구분되게 부여